

Módulo 2: Panorama general de herramientas y técnicas de digitalización según los distintos formatos





En esta clase veremos un panorama general de las herramientas y técnicas más habituales para la digitalización de archivos culturales. Para comenzar, algunas advertencias:

Sin dudas, cada obra y cada archivo es un mundo, por eso daremos apenas un panorama, lo que implica no profundizar en situaciones muy específicas que requieren un tratamiento especial (aunque vamos a trabajar en cada caso concreto en las actividades del curso). Pondremos atención entonces a los tipos de archivos más habituales: documentos impresos, obras plásticas, registros sonoros y audiovisuales.

La tecnología, por otro lado, ofrece una enorme variedad de herramientas y se desarrolla aceleradamente, por lo que no pretendemos abarcar todos los adelantos en este breve

panorama. Nos concentraremos en las tecnologías que pueden estar más al alcance de un proyecto de digitalización de modesto a mediano porte, con equipamiento posible de adquirir e incorporar por instituciones culturales con presupuesto y personal acotado.

Finalmente, la información que aquí presentamos puede ser discutida, y para eso está. Las técnicas propuestas tienen sus pros y sus contras; nos limitaremos a explicar cuáles son, para que cada persona sea quien, finalmente, con su experiencia, encuentre la mejor adaptación a su caso y necesidades. Si alguna de las recomendaciones que aquí damos resulta que no es aplicable, o se encuentra en la práctica una mejor solución, eso solamente significa que aquí no estamos escribiendo una biblia de la digitalización, sino solamente una guía inicial.

¿Qué es digitalizar?

Nos parece importante comenzar explicando aunque sea rápidamente qué es "lo digital", para recalcar que la digitalización es mucho más que convertir archivos físicos en algo incorpóreo para tenerlos en discos y acceder mediante computadoras.

En términos técnicos, la digitalización es un proceso por el cual el material de origen analógico se transforma al lenguaje binario de los *bits*: una combinación de ceros y unos que representan impulsos eléctricos. Al digitalizar, transformamos información de variado tipo y origen en una señal eléctrica que puede ser almacenada en diversos dispositivos y transmitida por redes, como una red de fibra óptica. ¿Pero qué implica todo esto?

En primer lugar: podemos guardar cada vez más cantidad de información en volúmenes físicos cada vez más reducidos; podemos transmitir esa información de manera prácticamente instantánea y llevarla a casi cualquier lugar del planeta. Esto significa que un archivo digitalizado puede ser consultado simultáneamente por muchas personas desde distintos lugares y que cada nueva consulta no implica el desgaste del objeto físico original.

Otra implicancia importante de la digitalización es que se produce una convergencia, ya que los distintos tipos de códigos y señales característicos de cada soporte -desde la cinta de un cassette hasta las letras impresas- se traducen en última instancia a un mismo tipo de codificación: el lenguaje digital. En consecuencia, los medios de almacenamiento y transmisión pueden ser los mismos para todos los tipos de archivos que se nos ocurra crear.

Esto significa que a su vez, los archivos digitales pueden ser transformados fácilmente a distintos formatos: con las herramientas adecuadas es posible desde el reconocimiento óptico de caracteres (OCR) que transforma la representación gráfica de cada letra a caracteres de texto digital, hasta la conversión automatizada de ese texto a voz para la lectura por parte de personas ciegas o de baja visión.

Pero hay más ventajas aún: el archivo digital puede guardar información acerca de sí mismo (metadatos), lo que facilita la búsqueda de información, y hace más simple su incorporación a otras colecciones manteniendo los datos de origen. Por ejemplo: podemos guardar en la colección digital de una biblioteca local una imagen obtenida de la biblioteca nacional, conservando todos los datos de catalogación: autor, año de edición, temas, etc.

Sin embargo, hay que tener en cuenta algunas limitaciones. En el proceso de digitalización la información analógica es traducida a bits, dando como resultado una representación imperfecta en la que siempre habrá ciertas pérdidas de información. Esta pérdida puede ser relevante o no, según para qué fines se digitaliza. Por ejemplo: una fotografía digital de resolución media a baja, puede ser una excelente representación de un texto escrito para el ojo humano, pero podría no servir para el reconocimiento óptico de caracteres, que requiere una imagen de alta resolución.

Otra limitación de los archivos digitales está dada por la obsolescencia de los formatos y de los programas que abren y modifican estos archivos. Por ejemplo, si almacenamos un libro en formato .DOC de Microsoft Word, es posible que en algún momento del futuro, si el fabricante de ese programa deja de desarrollarlo y cae en desuso, ese archivo ya no pueda ser abierto ni editado, y prácticamente lo perderíamos. Por eso es importante trabajar siempre con los denominados formatos abiertos, cuyas especificaciones son públicamente conocidas y responden a estándares que se van manteniendo con el consenso de organismos internacionales y participación de los desarrolladores de software. Para el caso citado, un formato abierto y estándar para documentos de texto sería .PDF o .ODF, por ejemplo.

Digitalización de documentos de texto e imagen

El tipo de documentos que encontramos en instituciones culturales como museos, archivos y bibliotecas, es claramente distinto del que se guarda en una empresa, un organismo burocrático o incluso en una casa. Estos documentos no interesan solamente por la información que contienen, sino que se los quiere preservar por el significado que tienen para la memoria colectiva. Así que no serán digitalizados únicamente para “hacer lugar” y poder deshacerse de una gran cantidad de papel, reduciendo el espacio de almacenamiento. Al digitalizar este material, casi siempre va a ser muy importante la preservación del original, por lo que se usarán procedimientos que conserven entero y en buenas condiciones cada ejemplar.

Las herramientas más usadas en este caso son los escáneres especiales para libros, que cuentan con cuna en forma de V en la que se coloca el libro y se lo abre en su ángulo de lectura natural, sin forzarlo. Pensemos que en un escáner común de “cama plana” hay que abrir el libro completamente y además apretarlo bajo la tapa del escáner, afectando la encuadernación. O pensemos en los escáneres de alimentación automática, similares a las fotocopiadoras, que requieren que las hojas estén sueltas y por lo tanto obligan a desencuadernar el libro.



Escáner Canon PIXMA MP830 con alimentación automática

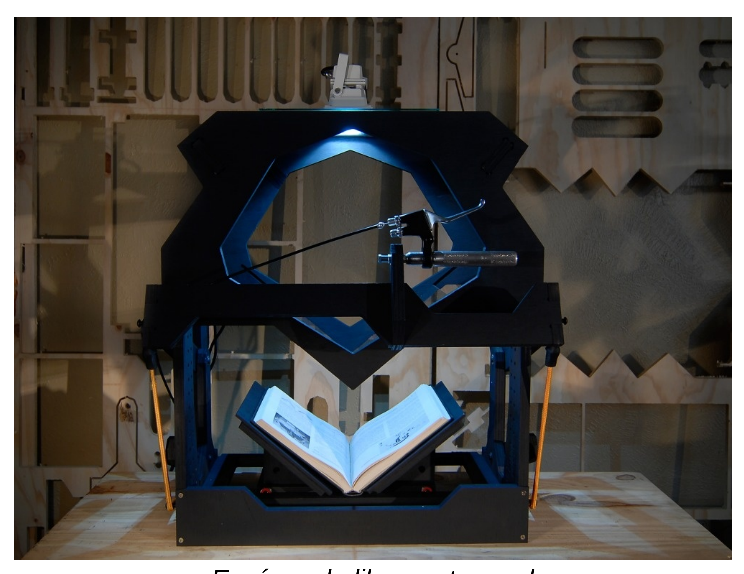


Escáner Plustek 4800 de cama plana



Escáner de libros industrial

http://www.book2net.net



Escáner de libros artesanal

http://diybookscanner.org

Los escáneres de libros pueden capturar digitalmente las hojas de los libros a gran velocidad utilizando cámaras de fotos. Por lo general producen archivos de imagen que se guardan en formato .TIFF, que son adecuados para hacer copias de preservación, y que pueden a su vez ser transformados para generar libros digitales amigables para el usuario. Con el trabajo de post-procesamiento que sigue a la captura se puede llegar a un .PDF o a un .EPUB, más adecuados para un lector que quiere consultar el archivo. El procedimiento que se usa para los libros también es aplicable a revistas, prensa y otro tipo de documentación, aunque estos pueden tener requerimientos especiales según el tamaño de los ejemplares, el tipo de encuadernación y la proporción entre texto e imagen.

En todos los casos, estos aparatos suelen ser bastante caros, pero también existen alternativas de bajo costo. El proyecto DIY Book Scanner facilita instrucciones para montar escáneres de libros económicos con materiales y tecnologías accesibles para cualquier institución cultural. En este video se puede ver una demostración de su funcionamiento.

Algunos recursos recomendados:

Comunidad de escáneres de libros Do It Yourself:
http://www.diybookscanner.org/es/index.html

Una guía clásica para la digitalización de materiales de bibliotecas es el manual de IFLA "Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos". Si bien es de 2002, muchas de las recomendaciones siguen vigentes http://www.ifla.org/files/assets/preservation-and-conservation/publications/digitization-projects-guidelines-es.pdf

Una guía más nueva pero mucho menos completa de IFLA: "Directrices para planificar la digitalización de colecciones de libros impresos antiguos y manuscritos"
http://www.ifla.org/files/assets/rare-books-and-manuscripts/rbms-guidelines/guidelines-for-planning-digitization-es.pdf

Proceso de digitalización en la Biblioteca Nacional de España. Es la guía de digitalización de dicha institución para su proyecto de Biblioteca Digital Hispánica:
http://www.bne.es/webdocs/Catalogos/ProcesoDigitalizacionBNE.pdf

Obras plásticas

Las obras de carácter plástico, como cuadros y esculturas, se digitalizan también mediante fotografías. Sin embargo, es un tipo de fotografía especial, que requiere de gran cuidado para lograr una excelente resolución y fidelidad respecto del original.



Frame del del video “How to photograph your art”, de Tyler Stalman y Jason Eng, publicado en Saatchi Online.

El procedimiento de fotografía digital común no es suficiente: hay que montar un set de fotografía especial. Se requiere una cámara que permita hacer ajustes manuales, un trípode para fijarla y fuentes de iluminación pareja para que la imagen quede homogénea. La cámara y los focos serán colocados de forma tal que no deformen la proporción del cuadro ni aparezcan brillos indeseados. Incluso puede ser necesarios filtros para evitar que la tonalidad de las fuentes de iluminación artificiales alteren el color.

Es recomendable almacenar copias de preservación de los archivos en la mejor calidad posible, sin compresión, o con un formato de compresión sin pérdida de calidad. Los más recomendables son .RAW, .TIFF o .PNG. Se debe evitar un exceso de edición para intentar “mejorar” la imagen, porque ésta puede sufrir pérdidas y alteraciones indeseadas. Se suele utilizar una carta de color al momento de capturar la imagen, que luego en la edición sirve de guía para el post-procesamiento, para que al momento de aplicar el retoque digital la imagen refleje lo más fielmente posible al original.

Algunos recursos recomendados:

Sobre materiales, accesorios y procedimientos para la digitalización de cuadros puede profundizarse en esta guía para hacer reproducciones fotográficas de obras de arte.

Una guía un poco más sencilla para comenzar a fotografiar obras de arte puede consultarse en la web de Ártica: http://www.articaonline.com/2012/07/como-fotografiar-nuestras-obras-de-arte-para-subirlas-a-internet/

Tutorial de digitalización de imagen de la Biblioteca de la Universidad de Cornell: http://preservationtutorial.library.cornell.edu/tutorial-spanish/tutorial_Spanish.pdf

Para profundizar en aspectos de preservación de la imagen digital, es interesante leer el artículo "La imagen digital, el valor de lo intangible", de Lino García Morales y Victoria Gutiérrez Colino.

Patrimonio sonoro y audiovisual

Los soportes para la reproducción de sonido y de imagen en movimiento han cambiado muy rápidamente en épocas relativamente recientes. Desde fines del siglo XIX hasta nuestros días, las tecnologías de grabación, almacenamiento y reproducción han avanzado de manera exponencial, pero la contracara de este proceso fue la rápida obsolescencia de los soportes y formatos, lo cual plantea un desafío para la preservación y, sobre todo, para el acceso.



Diversos tipos de soportes de imagen en movimiento: 1.- Cinta de 1 pulgada; 2.- Videocasetes de U-Matic ¾", estacionario y portátil; 3.- Betacam SP, Estacionario y portátil; 4.- Video8; 5.- DVCAM y miniDV; 6.- XDCAM). Fuente: http://www.andreadicastro.com/academia/CONFERENCIAS/Nueva_era_Video.html

Existe la ilusión de que siempre se va a poder inventar un nuevo soporte que aumente la calidad y prolongue la vida útil del registro, hasta llegar casi al "definitivo": el que conservará

incorruptible y siempre disponible el material. Pero el tiempo ha ido dejando obsoletos muchos soportes y sus correspondientes formatos y equipos de reproducción, por lo que la necesidad de trasladar el contenido a nuevos soportes se vuelve una tarea casi permanente. Incluso hoy, con la digitalización, el problema persiste y hasta puede acentuarse por la proliferación de nuevos soportes que requieren nuevas tecnologías. Pensemos en los archivos que fueron almacenados en CD desde los años 90. Los equipos de audio y computadoras desde hace un tiempo empiezan a venir sin bandeja lectora, porque los discos rígidos, las memorias de estado sólido y el almacenamiento en "la nube" están volviendo obsoleto al disco láser, que además demostró ser mucho más frágil y menos durable de lo prometido.

Este problema no es del todo solucionable, pero sí se pueden prever algunas cosas, como por ejemplo, asegurar copias digitales de preservación en la mayor calidad posible, en formatos abiertos y estándares, y contar además con planes de respaldo y migración para que los archivos no queden obsoletos por no poder ser reproducidos por las nuevas tecnologías.

Hay muchos procesos posibles de digitalización del patrimonio sonoro y audiovisual, porque hay muchos soportes diferentes, pero lo fundamental es contar con un equipo reproductor del soporte original que se pueda conectar a la computadora, la que debe tener una tarjeta de sonido y un software que gestione la digitalización. Por ejemplo: un cassette de audio se digitalizará conectando un equipo de música a la entrada de línea de la computadora que se encuentra a la vez conectada a la tarjeta de sonido interna de la misma. Se grabará el audio mediante un software de grabación de sonido, como Audacity, simplemente reproduciendo el cassette e indicando al programa que lo utilice como fuente de sonido.

El proceso de digitalización de video y audio analógicos implica tres pasos: muestreo, cuantificación y codificación. En este proceso se determina la fidelidad de imagen y sonido del archivo digital, por lo que las decisiones que se tomen en estos pasos determinarán la calidad de la reproducción del archivo y el tamaño del mismo. Una de las decisiones más importantes es qué códec utilizar para el almacenamiento, y aquí hay tantas respuestas como necesidades y objetivos tienen los archivos, de acuerdo al material a digitalizar. No tendrá los mismos requerimientos técnicos y de calidad la remasterización digital de una obra cinematográfica para su reestreno en salas, que la digitalización de un VHS familiar.

Pero por lo general, será necesario contar con una copia de preservación de la mayor calidad posible, y generar versiones que puedan ser transmitidas por streaming en Internet sin que requieran tanto espacio de almacenamiento y ancho de banda. Un códec recomendado para audio es FLAC, que es un formato de compresión sin pérdida de calidad, abierto y estándar. Los archivos se podrán difundir luego en otros formatos más comprimidos, aunque sufran pérdidas, como OGG o MP3 (este último muy común, pero no totalmente libre). En video hay muchas opciones, pero en general para el almacenamiento de preservación se recomiendan AVI, MXF y MOV, sin compresión, mientras que para el acceso por Internet puede cumplirse con un archivo MPEG, más comprimido.

Algunos recursos recomendados:

Aspectos técnicos de la digitalización de fondos audiovisuales, por David González-Ruiz, Miquel Térmens y Mireia Ribera:
http://www.elprofesionaldelainformacion.com/contenidos/2012/septiembre/12.pdf

Introducción a los códecs de archivos sonoros y audiovisuales y 10 recomendaciones para seleccionar y gestionar códecs, por Chris Lacinak: http://www.avpreserve.com/wp-content/uploads/2014/10/PrimerForCodecs_spanish.pdf

Cómo digitalizar VHS, casetes de audio y discos de vinilo viejos para preservar la memoria:
http://www.lanacion.com.ar/838922-como-digitalizar-vhs-casetes-de-audio-y-discos-de-vinilo-viejos-para-preservar-la-memoria

Recomendaciones generales

Independientemente del tipo de obras que se digitalizan y de la técnica utilizada, hay algunas cuestiones generales de metodología que se aplican a todos los procedimientos:

- **Establecer un protocolo con las especificaciones técnicas de la digitalización.** Por ejemplo, procedimientos de manipulación de los originales, etapas de digitalización, formato, calidad y niveles de compresión de los archivos digitales tanto para las copias de preservación como las de difusión, parámetros del post-procesamiento, estándares de metadatos, etc. A modo de ejemplo, ver el protocolo de la Biblioteca Nacional de Chile para la digitalización de imágenes.
- **Llevar controles y registros** en una planilla para registrar y controlar la marcha del proceso y sus etapas. Como mínimo, debe incluir el número de identificación de cada material, persona que digitaliza, persona que post-procesa, estado del proceso, fechas de inicio y fin de la digitalización y observaciones relevantes.
- **Trabajar en todas las etapas que sea posible con software libre y formatos abiertos** para no quedar atados a productos tecnológicos que imponen restricciones y nos hacen dependientes de un único producto para recuperar, reproducir y editar los archivos.
- A diferencia de lo que recomiendan muchos manuales de digitalización de bibliotecas, nosotros creemos que es importante **no colocar marcas de agua en los archivos digitales**. Las marcas de agua alteran el material original y entorpecen su uso y circulación. Toda la información añadida a la obra que sea necesario comunicar acerca de la institución que realiza la digitalización, puede incorporarse en los metadatos.
- **Desarrollar planes para el respaldo y eventualmente, la migración de archivos**, como parte de una política de preservación digital. Se recomienda partir de las directrices para la preservación del patrimonio digital de la UNESCO.